ماهر أسعد بكر

الذكاء الاصطناعي والمعلومات المضللة

ماهر أسعد بكر

الذكاء الاصطناعي والمعلومات المضللة

ISBN: 9798215050392

COVER IMAGE MADE USING: HTTPS://WWW.FREEPIK.COM

محتويات

خلاصة

في العالم الرقمي اليوم، تتدفق المعلومات بحرية وبلا نهاية عبر المنصات. ورغم أن هذا الاتصال مكّن من نشر الأفكار، فقد مكّن أيضاً من نشر الادعاءات المضللة التي يمكن أن تؤدي إلى تآكل الحقيقة والثقة، مع ظهور تهديدٍ جديدٍ متمثلٍ بالمد المتصاعد للذكاء الاصطناعي، حيث تسمح النماذج التوليدية بإنتاج محتوى اصطناعي ولكنه متطور. ومثل الفيضان، تهدد المعلومات المضللة التي يولدها الذكاء الاصطناعي بإغراق أسس المجتمع المستنير تحت طوفان من الأكاذيب.

علينا أن نعترف بأن هذا الفيضان لم يكن غير متوقع، بل كان نتيجة حتمية للتطوير والانتشار غير المبالي دون مسؤوليةٍ أو حكمةٍ كافية.

لقد فشل مبتكرو هذه الأدوات القوية في الأخذ في الاعتبار التطبيقات الأكثر قتامة والحاجة إلى إنشاء حواجز حماية ضد إساءة الاستخدام. والآن ينتشر الخطر إلى حدٍ كبيرٍ دون رادع، و تكافح المنصات لوقف التيار، ويترك الأفراد يتخبطون، ورغم أن التنظيم والاعتدال لهما أدوار، فإن الحل يكمن، و بقدرٍ أعمق من ذلك، في الاستيلاء على ملكية تقدمنا التكنولوجي وإعطاء الأولوية للنزاهة على النفعية.

ولن ينحسر هذا المد من خلال رد الفعل وحده، بل يتطلب التحول نحو المسؤولية الاستباقية. يجب أن يدرك المطورون واجبهم الأخلاقي في تصور إساءة الاستخدام وتنفيذ الضمانات التي لا تؤثر على الوظائف فقط ولكنها تدعم الموثوقية. يجب أن تجد المنصات الشفافية وأن تقوم بتثقيف المستخدمين لتعزيز التفكير النقدي ضد التلاعب. وللأفراد أيضاً دورٌ في محو الأمية الإعلامية والرغبة في إعادة النظر في المفاهيم المسبقة في مواجهة الأدلة المتناقضة. يجب علينا

أن ندعم أسسنا بالحق والحكمة قبل العاصفة القادمة.

تهدف هذه الأطروحة إلى مسح مياه الفيضانات المتزايدة وتقييم دفاعاتنا، وتحدد مفهومي المعلومات الخاطئة والمعلومات المضللة، مع التمييز بين الباطل حسب النية. سوف تدرس كيفية تمكين الذكاء الاصطناعي التوليدي من إنتاج كميات كبيرة من المحتوى الاصطناعي، بدءاً من التزييف العميق deepfakes وحتى النصوص المقنعة، والمواد الواقعية والمضللة التي يسهلها ذلك. سيتم استكشاف طرق الكشف، مع الأخذ في الاعتبار التقنيات بدءاً من التحقق من الحقائق وحتى تحليل المصدر. سيتم أيضاً تقييم سياسات النظام الأساسي والمسؤوليات الفردية.

ومن خلال التحليل الدقيق، يسعى هذا العمل إلى إضفاء الوضوح

والمسؤولية على هذه القضية. وهي لا تهدف إلى إنكار التقدم التكنولوجي ولا استبعاد المخاطر، بل تهدف إلى السير في مسار متوازن بين الاثنين. وفي نهاية المطاف، فإنها تدعونا إلى الارتفاع فوق ردود الفعل وتبني البصيرة، وإعطاء الأولوية للنزاهة على الراحة في إنشاء المعلومات واستهلاكها. إن مدّ الكذب ينمو بسرعة، ولكن بالحكمة والشجاعة، يمكننا أن نبني حصوناً من الحقيقة لمقاومة الطوفان القادم. ومن خلال فهم التحديات والتعاون بشكلٍ بناء، يستطيع مجتمعنا أن يخرج من هذه التجربة وقد تم تعزيز أسسه بدلاً من تآكلها. لقد حان الوقت للاستعداد، قبل الطوفان.

كيف يولد الذكاء الاصطناعي معلومات مضللة

إن شبح الفوضى التكنولوجية يلوح في الأفق بشكل خطير، ويهدد بقلب النظام الهش الذي يشكل الركيزة الأساسية لسلامتنا العقلية. هذه الإنتروبيا، الأقرب للتفاؤل الساذج، تنبع لأنبياء السيليكون الذين يتمسكون بإخلاص بإيديولوجية التقدم، في حين يظلون جاهلين بالظلال المدمرة التي تلقيها إبداعاتهم على الواقع. و هنا نحن نتحدث عن صعود الذكاء الاصطناعي التوليدي، وقدرته على استحضار المعلومات الخاطئة والتزييف الذي قد يمزق نسيج الحقيقة نفسها.

من الخطأ أن نعتقد أن هذه الأدوات بسيطة, بل في الحقيقة إنها

أدواتٌ قوية، وقادرةٌ على محاكاة الإبداع البشري على نطاق يفوق الخيال. يمكن لنماذج مثل 3-GPT إنتاج تدفقات متتالية من النصوص المتماسكة، كما لو أن الأفكار تم انتزاعها بالجملة من فراغٍ لا حدود له. وبالمثل، تستحضر شبكات GAN صور الوجوه والمشاهد الواقعية من الضوضاء، ومع تقنية التزييف العميق deepfakes، تصبح كل الثقة في أدلة الفيديو موضع شك، حيث تعمل الشبكات العصبية على تحريك الوجوه وخلق الأصوات بسلاسة.

إن استخدام لهذه الأدوات دون انتباه، يعني اللعب بقوىً قد تفلت من السيطرة بسهولة. ما الذي سيحمينا من طوفان الوسائط المزيفة الذي قد ينتشر؟ يمكن للجهات الفاعلة السيئة أن تزرع الفتنة من خلال نشر دعايةٍ ماكرةٍ مصنوعة بالكامل من الأكواد. قد تؤدي المراجعات الاصطناعية إلى الافتراء والتملق. الهويات المزورة يمكن أن تتسلل وتخدع دون كلل. إن حجم الإنتاج سوف يتجاوز بكثير

التحقيقات المتثاقلة التي يقوم بها مدققو الحقائق، الذين يطغى عليهم الوفرة التكنولوجية.

وماذا ستصبح الحقيقة في مثل هذا العالم؟ إن الحقائق، وهي حجر الأساس للعقلانية، تخاطر بالذوبان في فوضى الخداع. يتمرد العقل ضد هذه النسبية غير المقيدة. ومع الحرمانهم من المراسي الآمنة وسط اضطراب الخطاب الاصطناعي، يصبح البحث عن الحقيقة مسألة إيمان وليس سبباً. قد يتراجع البعض إلى المؤامرة الوهمية، بينما يخضع البعض الآخر للسجود أمام أقوال التكنولوجيا.

ما نحتاج إليه هو تجديد الثقة بالواقع أمام حواسنا، مصحوباً بالمزيد من البصيرة لاكتشاف الخداع الرقمي الذي قد يحيط بنا. يمكن للتكنولوجيا أن تضخم الروح الإنسانية، ولكنها قد تؤدي أيضاً إلى

ظهور شياطيننا الأدنى إذا تركتها التقاليد الحكيمة دون رادع. يجب علينا توجيه هذه الأدوات نحو الإبداع وقول الحقيقة، وتمييز التطبيقات الصالحة بينما نعارض كل الإنتروبيا والخداع. وربما يمكننا بعد ذلك بناء نظام لتهدئة العاصفة التكنولوجية القادمة. ولكن يتعين علينا أن نتحرك بسرعة، بشجاعة وعناية، خشية أن ينهار الفيضان دون عوائق.

تقدم الأحداث الأخيرة حكايات تحذيرية عما قد يحدث إذا استمرت الاتجاهات الحالية بلا هوادة. في عام 2018، ظهر مقطع فيديو يظهر سياسيةً بارزةً وهي تتلعثم في كلماتها، مما أدى لتعزيز الشك في قدراتها. كشف تحليل الطب الشرعي عن تغييرٍ رقمي متطور، وقد كان نذيراً لـ " التزييف العميق " deepfakes القادم.

وشهد ذلك العام أيضاً تلاعباً فظاً ولكن مثيراً للقلق بتصريحات القادة السابقين، مما أدى إلى زرع بذور عدم الثقة. وفي عام 2020، استهدفت عمليات خداع أكثر تقدماً العملية الديمقراطية، في إشارة إلى تكتيكات قد تقوض أسس الحقيقة والمجتمع. أظهرت هذه الغزوات المبكرة أن الباب بات مفتوحاً أمام عالمٍ لما بعد الواقع، حيث تصبح الأصالة نفسها عرضةً للمراجعة.

تعمل نماذج الذكاء الاصطناعي المتخصصة الآن على توليد كميات كبيرة من النصوص المكتوبة آلياً والتي لا يمكن تمييزها تقريباً عن تلك التي أنشأها الإنسان، مما يسمح بإنتاج كميات كبيرة من الأخبار المزيفة والمراجعات والمشاركات الشخصية حول أي موضوع. وفي الوقت نفسه، تقوم خوارزميات الرؤية الحاسوبية بتجميع صور واقعية لأشخاص وأحداث غير حقيقية لم تكن موجودة من قبل.

يستمر " التزييف العميق " Deepfake في التحسن، و يتم إدراج أي وجه بسلاسة في الأفلام باستخدام تعبيرات دقيقة و مطابقة حركات الشفاه. إن العائق أمام التلاعب المتطور سيكون قريباً هو المهارة وليس التكنولوجيا مع انتشار هذه القدرات على نطاق أوسع. وقد يستخدمها أولئك الذين يميلون إلى ذلك لتحقيق أهدافٍ سياسيةٍ أو ماليةٍ أو أيديولوجيةٍ لنشر الأكاذيب وتقويض المعارضين.

من المحتمل أيضاً أن يتم جني الأموال من خلال استغلال مخاوف الجمهور من خلال العلاجات المزيفة أو التلاعب بالأسهم. ومع إضفاء المزيد من الديمقراطية على الوسائل، تزداد صعوبة مراقبة الاستخدام الضار بشكل كبير. نحن نتجه نحو مشهد ما بعد الواقع حيث يتطلب تمييز الواقع التدقيق، ولا يمكن قبول أي ادعاء دون التحقيق في الاستجواب.

اكتشاف المعلومات الخاطئة الناتجة عن الذكاء الاصطناعي

مع استمرار نماذج الذكاء الاصطناعي التوليدية في تقدمها بلا هوادة، تصبح مهمة اكتشاف المخرجات الاصطناعية التي تنتجها صعبةً بشكلٍ متزايد. ومع ذلك، هناك أمل في الأفق، حيث أظهرت العديد من التقنيات نتائج واعدة عند دمجها كجزء من استراتيجيةٍ شاملة.

أولا وقبل كل شيء، أثبت اكتشاف التعلم الآلي أنه أداةٌ قيمةٌ في مكافحة المعلومات الخاطئة التي يولدها الذكاء الاصطناعي. كرّس الباحثون جهودهم لتدريب نماذج التعلم الآلي المصممة خصيصاً

لاكتشاف التزييف العميق والأشكال الأخرى من الوسائط التي تم التلاعب بها. يمكن لهذه النماذج تحليل التناقضات الدقيقة التي لا يمكن رؤيتها بالعين البشرية. على سبيل المثال، عندما يتعلق الأمر بمقاطع الفيديو المزيفة ، تقوم هذه النماذج بفحص الطريقة التي يتم بها مزج الوجوه بسلاسة، والبحث عن أنماط غير طبيعية تعمل كمؤشرات للتلاعب. وبالمثل، عندما يتعلق الأمر بنماذج النص التوليدية، فإنها قد تقدم حالات شاذة إحصائية أو تناقضات في بناء الجملة مما يكشف عن أصولها الاصطناعية.

مع استمرار تطور مجال التقنيات التوليدية، يجب أن تواكب نماذج الكشف البحث المستمر وجمع البيانات. أظهرت الأساليب متعددة الوسائط التي تحلل العناصر المرئية والنصية أداءً محسناً مقارنة بالتحليل من النوع الواحد. ومع ذلك، من المهم أن نعترف بأن هذه المعركة هي أقرب إلى "سباق التسلح". ومع تحسن النماذج

التوليدية، فإنها تفرض تحدياتٍ جديدةٍ على أجهزة الكشف الموجودة، مما يستلزم إعادة التدريب والتكيف المستمر. للبقاء في صدارة اللعبة، يعد التعاون بين الباحثين في مجال سلامة الذكاء الاصطناعي وبناة النماذج التوليدية أمراً بالغ الأهمية.

بالانتقال إلى ما هو أبعد من التعلم الآلي، يظهر تحليل البيانات الوصفية كأداة قوية أخرى في الترسانة ضد المعلومات الخاطئة الناتجة عن الذكاء الاصطناعي. تشير البيانات الوصفية، في هذا السياق، إلى المعلومات السياقية المرتبطة بجزءٍ من المحتوى. تأتي الوسائط الحقيقية عادةً مع بيانات التعريف التي تتضمن توقيعات الجهاز وبيانات الموقع وتاريخ التحرير. غالباً ما تكون هذه المعلومات المهمة غائبة في المحتوى الاصطناعي. علاوة على ذلك، قد تفتقر صفحات الويب التي تحتوي على نص تم إنشاؤه بواسطة الذكاء الاصطناعي إلى مراجع للمصادر والمؤلفين والطوابع الزمنية.

لتحقيق أقصى استفادة من البيانات الوصفية، تم تطوير أدوات مثل Project Provenance لتحليل هذه المعلومات وتحديد احتمالية التلاعب بعنصر معين. ومع ذلك، من الضروري الاعتراف بإمكانية تزوير البيانات الوصفية أو إزالتها في محاولة لتجنب الكشف. ولمواجهة ذلك، يمكن أن تساعد البيانات الوصفية المرجعية مع تقارير موثوقة من مصادر موثوقة في تحديد المصدر. علاوة على ذلك، تساهم الجهود المنسقة للتحقق من الحقائق في تعزيز عملية التحقق من خلال السماح بإجراء المقارنة عبر تحليلات مستقلة متعددة.

وفي حين تلعب الأدوات التقنية دوراً حاسماً، فإن الحكم البشري يظل الأساس ولا يمكن استبداله في المعركة ضد المعلومات المضللة التي يولدها الذكاء الاصطناعي. على الرغم من التقدم الملحوظ في الذكاء

الاصطناعي، فإن هذه الأنظمة قد لا تلتقط بشكل مثالي الاختلافات الدقيقة بين المحتوى البشري والمحتوى الناتج عن الآلة. وبالتالي، فإن إشراك الجهات الفاعلة البشرية يصبح أمراً لا غنى عنه. يلعب الصحفيون والمحققون ومشرفو المنصات أدواراً حيوية في تأكيد النتائج الفنية وتطبيق الفهم السياقي.

أحد الأساليب الفعالة لتسخير الحكم البشري هو من خلال مبادرات المراجعة الجماعية مثل ClaimReview، والتي تستفيد من الذكاء الجماعي للجمهور لتقييم المطالبات. ومن خلال إشراك مجموعة متنوعة من الأفراد في عملية التقييم، يمكننا الاستفادة من وجهات نظرهم وخبراتهم الفريدة. وبالمثل، يمكن للمنصات دمج سير العمل البشري داخل الحلقة، مما يسمح بتحديد المحتوى الاصطناعي المحتمل ووضع علامةٍ عليه لمراجعته مجتمعياً. ومن خلال إشراك المجتمع الأوسع، فإننا نستفيد من الحكمة الجماعية ونستفيد من

مجموعةٍ واسعةٍ من الأفكار. بالإضافة إلى ذلك، فإن تثقيف الأفراد بشأن موثوقية المصدر وتزويدهم بالمهارات اللازمة لاكتشاف التناقضات المنطقية يمكن أن يعزز قدرتهم على تقييم المعلومات بشكلٍ مدروس.

تقترح الدراسات تضمين العلامات المائية ضمن النماذج التوليدية، وربط المحتوى الاصطناعي بشكلٍ جوهريٍ بأصوله، كما أن الفحص الفني يشير إلى الوسائط المشبوهة في الوقت الفعلي، وتوفر عمليات التحقق من البيانات الوصفية/المصدر سياقاً بالغ الأهمية.

تستمر دقة الكشف في التقدم مقابل تحسين التهديدات. يمكن للجهود المتواصلة عبر الأبحاث والسياسات والتكنولوجيا والمجتمع مكافحة الخداع من خلال الذكاء الاصطناعي. وبالحكمة والتعاون بين جميع الأطراف الفاعلة، يمكننا أن نتمسك بالحقيقة حتى في مواجهة

التهديدات التوليدية الناشئة.

إن النجاحات الأولية في الدفاع عن الحقيقة من خلال التعاون تخلق الأمل. والآن يجب أن تصبح اليقظة والاستعداد من الأولويات المجتمعية، وبالاضافة لتعزيز الأدوات التقنية بمجتمعاتٍ مطلعةٍ قادرةٍ على الصمود في وجه التلاعب. وباسترشادها بهذه الطريقة فإن القوى الناشئة قد تخدم العدالة بدلاً من تقويضها.

استجابات المنصات للمشكلة

في مشهد يسود فيه نشر المعلومات عبر الإنترنت، تبرز منصات وسائل التواصل الاجتماعي باعتبارها القنوات الأساسية، وتمارس تأثيراً هائلاً على انتشار المعلومات الخاطئة التي يولدها الذكاء الاصطناعي. ومن الضروري تحليل السياسات الحالية وجهود الشفافية التي تبذلها المنصات الرئيسية لاستنباط مناهج معززة يمكنها حماية الحقيقة بشكلٍ فعالٍ داخل المجال العام الرقمي.

الإشراف على المحتوى

في حين قطعت المنصات خطواتٍ كبيرة في تحديث معايير مجتمعها

لتحظر صراحة التزييف العميق وأشكال معينة من الوسائط الاصطناعية، فإن عملية الاعتدال تظل في الغالب تفاعلية. بحلول الوقت الذي يتم فيه وضع علامة على المحتوى الإشكالي وإزالته، ربما يكون قد تم نشره بالفعل على نطاقٍ واسع، مما يترك تأثيراً دائماً على الإدراك العام. ولمعالجة هذه المشكلة، يمكن أن يكون الفحص المسبق للحسابات التي لها تاريخٌ من الانتهاكات بمثابة إجراء وقائي، مما يؤدي إلى إبطاء انتشار المعلومات الخاطئة.

تلعب أدوات الكشف التلقائي أيضاً دوراً محورياً في دعم جهود الإشراف على المحتوى من خلال وضع علامة على المحتوى الذي قد يسبب مشاكل. ومع ذلك فإن الفروق الدقيقة في السياق تجعل من الضروري إشراك المراجعين البشريين في الحلقة. يجب أن تحقق المنصات توازناً دقيقاً بين الحفاظ على حرية التعبير ودعم سلامة المعلومات من خلال إعطاء الأولوية للعمليات الشفافة والمتسقة

على حساب التنفيذ الخوارزمي البحت.

تقارير الشفافية

لفهم مدى فعالية جهود الإشراف على محتوى المنصات، من الضروري أن تكون هناك شفافية فيما يتعلق بانتشار المحتوى الاصطناعي ومعدلات إزالته. وعلى الرغم من أن شركات مثل فيسبوك ويوتيوب تقدم تقارير منتظمة تسلط الضوء على هذه الجوانب، إلا أن الافتقار إلى الاتساق المنهجي يعيق المقارنة المباشرة. ومن شأن إنشاء مقاييس موحدة ومدققة من قبل طرف ثالث أن يعزز المساءلة ويسهل تحديد أولويات البحث في المجال التقني.

ويجب أن تكون المنصات أيضاً مستعدة للكشف عن التقنيات

المستخدمة للكشف عن المحتوى الاصطناعي، دون المساس بأساليب الملكية. تعمل المناقشة المفتوحة التي تحيط بقدرات تقنيات الكشف هذه وقيودها على تمكين المستخدمين وتدعو إلى الحصول على رؤى قيمة من الخبراء الخارجيين. تعمل الشفافية كأساس لبناء الثقة، مما يؤكد للمستخدمين أن المنصات تتعامل بجدية مع التهديدات الناشئة.

التعليم والتوعية

وفي حين تهدف الأدوات والسياسات التقنية إلى الحد من جانب العرض في مشكلة المعلومات المضللة، فمن الضروري الاعتراف بدور الطلب في دفع انتشارها. يمكن أن تكون الجهود التعاونية بين المنصات والمنظمات الخارجية مفيدةً في تعزيز المعرفة الإعلامية ومهارات التفكير النقدي لدى المستخدمين، وتزويد الجمهور بالقدرة

على تمييز المطالبات وتقييمها بدلاً من استهلاك المحتوى بشكلٍ سلبي.

حملات التوعية، مثل حملة "فكر قبل المشاركة"، تشكل دوراً مهماً في تحذير المستخدمين من المخاطر المرتبطة بالمحتوى الناتج عن الذكاء الاصطناعي. توفر أدوات التحقق من الحقائق المتكاملة إمكانية التحقق على المنصة من القصص المشبوهة، مما يمكّن المستخدمين من اتخاذ قرارات مستنيرة.

يساعد وضع العلامات الواضحة على المحتوى الاصطناعي في تحديد التوقعات وتعزيز فهم أنه قد تم إنشاؤه بشكل مصطنع. تعمل المجتمعات المطلعة على تعزيز ثقافة النزاهة في المناقشات عبر الإنترنت، لتكون بمثابة حصنٍ ضد انتشار المعلومات المضللة.

أنظمة

وفي الحالات التي تكون فيها الإجراءات التطوعية غير كافية، تصبح التدابير التنظيمية ضرورية لضمان المساءلة والمسؤولية. وقد تم بالفعل وضع سوابق مع القوانين التي تتناول الإعلانات السياسية والتزييف العميق، مما يسلط الضوء على الحاجة إلى الرقابة في هذه المجالات. وينبغي للمنصات أن تشارك بشكل استباقي مع صانعي السياسات والباحثين المستقلين لتطوير أطرٍ متوازنةٍ وقائمةٍ على المخاطر لتطوير واستخدام النماذج التوليدية. ويعزز التعاون الدولي الجهود الجماعية الرامية إلى مكافحة انتشار مشكلة المعلومات المضللة المعولمة.

تتطلب حماية الحقيقة في المجال العام الرقمي اتباع نهجٍ شاملٍ يشمل الإشراف على المحتوى، وتقارير الشفافية، والتعليم والتوعية،

والتنظيم. يجب أن تتبنى المنصات تدابير استباقية لمنع الانتشار الفيروسي للمعلومات الخاطئة مع الحفاظ على الشفافية في أفعالها وتعزيز ثقافة التفكير النقدي بين المستخدمين. يعد التعاون بين المنصات والمنظمات الخارجية وصناع السياسات والباحثين أمراً ضرورياً للتصدي بفعالية للتحديات التي تفرضها المعلومات الخاطئة الناتجة عن الذكاء الاصطناعي. ومن خلال تحصين دفاعاتنا والحفاظ على سلامة المعلومات، يمكننا أن نسعى جاهدين نحو نظام بيئيٍ أكثر صدقاً واستنارة عبر الإنترنت.

دراسة حالة

في هذا الفحص المتعمق، نستكشف مثالاً حقيقياً لمحتوى التزييف العميق الناتج عن الذكاء الاصطناعي والمنتشر عبر المشهد عبر الإنترنت، بهدف الكشف عن التحديات والدروس التي يمكن استخلاصها من هذه الحالة بالذات. وقع هذا الحدث في يونيو 2020 عندما ظهر مقطع فيديو مزيف على وسائل التواصل الاجتماعي، يُزعم أنه يعرض شخصيةً سياسيةً بارزةً تدلي بتصريحات مثيرة للجدل. انتشر الفيديو بشكلٍ سريع، حيث حصد أكثر من مليون مشاهدة في يوم واحد حيث قام المستخدمون المطمئنون بمشاركته دون قصد داخل شبكاتهم.

وباستخدام تقنياتٍ متطورةٍ تعتمد على التعلم العميق وتبديل الوجه، قدم الفيديو سيناريو مقنعاً ولكنه ملفق. ومع ذلك، بعد

التحليل الدقيق لكل إطار على حدة frame-by-frame، تمكن المراقبون الأذكياء من تمييز التناقضات الدقيقة في تعبيرات الوجه وحركات الشفاه التي تم التلاعب بها، مما كشف في النهاية عن طبيعته الاصطناعية. لسوء الحظ، بالنسبة للمشاهد العادي الذي يشاهد المقطع القصير خارج السياق، فقد بدا حقيقياً تماماً. تشير التقارير إلى أن الفيديو ظهر في البداية على حساب مجهول على إحدى منصات مشاركة الفيديو قبل أن ينتشر عبر منصات أخرى مختلفة.

فشل موقع مشاركة الفيديو الذي تم نشره فيه لأول مرة في اكتشاف وإزالة التزييف العميق لأكثر من 12 ساعة، مما سمح بالنشر الأولي على نطاق واسع. كما شهدت شبكات التواصل الاجتماعي الكبرى، مثل فيسبوك وتويتر، تأخيرات أو تناقضات في جهود الإزالة، حيث استمرت نسخ مكررة من الفيديو في الانتشار عبر عناوين URL بديلة

أو أعيد نشرها من حسابات دولية.

على الرغم من أن منظمات التحقق من الحقائق فضحت الفيديو بسرعة، إلا أنها كافحت لمواكبة الانتشار السريع للمحتوى. وكشف تحليل البيانات الوصفية أنه خلال فترة الـ 24 ساعة الأولى فقط، تم تنزيل الفيديو المزيف وإعادة نشره أكثر من 300 ألف مرة. ومن المثير للقلق أن الدراسات الاستقصائية أشارت إلى أن الفيديو استمر في ممارسة التأثير على تصورات المشاهدين، مع احتفاظ العديد من الأفراد بإيمانهم بالادعاءات الحرفية التي قدمها الفيديو بعد أسابيع، على الرغم من فضح الزيف لاحقاً .

تؤكد دراسة الحالة هذه على العديد من التحديات المهمة المرتبطة بالمعلومات الخاطئة الناتجة عن الذكاء الاصطناعي. فأولاً وقبل كل

شيء، يشكل الانتشار الأولي السريع لمثل هذا المحتوى قبل تنفيذ تدابير الكشف الفعالة مشكلةً كبيرة. علاوة على ذلك، فإن الانتشار الفيروسي عبر الأنظمة الأساسية للتزييف العميق يزيد من مدى وصولها وتأثيرها. إن التناقضات في استجابات المنصة، كما لوحظ في هذه الحالة، تزيد من تعقيد المشكلة، مما يستلزم بذل المزيد من جهود الاعتدال المنسقة. وأخيرا، فإن التأثير المستمر لمثل هذا المحتوى على تصورات المشاهدين، حتى بعد محاولات التحقق من الحقائق، يسلط الضوء على الحاجة إلى تقنيات كشف محسنة، وسياسات شفافة، ومبادراتٍ تعليميةٍ تعزز الثقافة الإعلامية الهامة.

ومن خلال الخوض في حالات واقعية مثل هذه، يمكننا الحصول على رؤىً قيمةً حول التحديات التي تفرضها المعلومات الخاطئة التي يولدها الذكاء الاصطناعي وتطوير استراتيجياتٍ أكثر فعاليةٍ لحماية الحقيقة في مواجهة التهديدات التوليدية الناشئة. وينبغي أن تشمل

هذه الاستراتيجيات تقنيات الكشف المحسنة، وجهود الإشراف التعاونية، وسياسات المنصات الشفافة، والمبادرات التعليمية التي تعزز الثقافة الإعلامية الهامة. ومن خلال الفهم الشامل لهذه القضايا المعقدة، يمكننا العمل على إيجاد مشهد رقمي أكثر مرونةً واستنارة.

خاتمة

تعمق هذا الطرح الشامل في الصعود المثير للقلق للمعلومات الخاطئة الناتجة عن الذكاء الاصطناعي وتأثيرها الضار على تآكل الحقيقة في العالم الرقمي. من خلال التدقيق في التقنيات التوليدية المختلفة المستخدمة، وطرق الكشف المستخدمة، والاستجابات من المنصات والهيئات التنظيمية، نهدف إلى فهم التحديات والفرص التي تكمن في حماية سلامة المعلومات في المجال العام الرقمي.

تسلط النتائج الرئيسية لهذا البحث الضوء على جوانب مهمةٍ من القضية المطروحة:

لقد أطلق ظهور الذكاء الاصطناعي التوليدي الحديث، وخاصة النماذج العصبية الكبيرة، العنان لموجة من المحتوى الاصطناعي المتطور والمضلل، بما في ذلك المحتوى المزيف العميق والمراجعات المزيفة. يؤدي هذا الانتشار للمعلومات الخاطئة الناتجة عن الذكاء الاصطناعي إلى تضخيم انتشار الروايات الكاذبة ويقوض مصداقية المعلومات الحقيقية.

على الرغم من أن طرق الكشف ذات قيمة، إلا أنها تظل غير كاملة بسبب التقدم المستمر في الذكاء الاصطناعي التوليدي والقيود المفروضة على البيانات الوصفية المتاحة. ومع ذلك، فإن الجهود المنسقة التي تجمع بين تقنيات التعلم الآلي، وتحليل المصدر، والحكم البشري تظهر الوعد الأكبر في مكافحة المعلومات الخاطئة التي يولدها الذكاء الاصطناعي.

وبينما بذلت المنصات جهوداً لتعزيز سياساتها، هناك حاجة ملحة للشفافية المتسقة والتعليم الذي يعطي الأولوية لرفاهية المجتمع على مجرد التدابير الرجعية. إن تحقيق التوازن بين حرية التعبير والحفاظ على النزاهة أمرٌ بالغ الأهمية في البحث عن الحقيقة.

يلعب التعاون التنظيمي دوراً حيوياً في تحديد المسؤوليات الأساسية مع التطور السريع للتكنولوجيا. ومن خلال تعزيز التعاون وتجريب السياسات المنفتحة، يمكن للجهات التنظيمية المساهمة في تطوير أطرٍ متوازنةٍ تعالج التحديات التي تفرضها المعلومات الخاطئة التي يولدها الذكاء الاصطناعي.

إن المضي قدماً والتخفيف الفعال من تأثير المعلومات الخاطئة الناتجة عن الذكاء الاصطناعي يتطلب التزاماً مستداماً من العديد من أصحاب المصلحة. يجب على الباحثين مواصلة جهودهم في تطوير تقنيات كشف قوية مع المشاركة بنشاط مع صناع السياسات لاتخاذ قراراتٍ سياسيةٍ سليمة. يجب على المنصات والمطورين إعطاء الأولوية للشفافية بشكل استباقي ودمج الضمانات التي تدعم النزاهة. يلعب المعلمون ومجموعات المجتمع دوراً حاسماً في تعزيز مهارات التفكير النقدي وتعزيز الحكمة للحماية من التلاعب. علاوة على ذلك، يتحمل الأفراد أنفسهم مسؤولية التقييم المدروس للادعاءات ودعم المناقشات القائمة على الحقائق. ومن خلال تعزيز أهمية الحقيقة من خلال الجهود التعاونية بدلاً من الاستجابات الرجعية، يصبح بوسعنا تنمية أنظمة معلوماتية مرنةٍ قادرةٍ على تحمل التحديات الناشئة للعمليات الديمقراطية.

إن معالجة هذه المشكلة المعقدة تتطلب حلولاً شاملة وروح التقدم المشترك التي تتجاوز المصالح. ومع الاجتهاد الذي لا يتزعزع والالتزام الثابت بالحقيقة، لا ينبغي تقويض وعد الذكاء الاصطناعي بسبب مخاطره المحتملة. وبدلاً من ذلك، من خلال اجتياز هذه المحاكمة مع احترامٍ عميقٍ للواقع والعدالة، لدينا الفرصة للخروج أقوى، مع أسسنا الراسخة في السعي وراء الحقيقة.

اقتباسات

1. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6).

2. Boididou, C., Andreadou, K., Papadopoulos, S., & Drosatos, G. (2021). Verifying the authenticity of images using deep learning. IEEE Access, 9, 77938-77949.

3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

4. Gehrmann, S., Strobelt, H., & Rush, A. M. (2021). GLTR:

Statistical detection and visualization of generated text. In Proceedings of the 2021 ACM Conference on Fairness, .(358-ccountability, and Transparency (pp. 349A

Major, A., & -Bender, E. M., Gebru, T., McMillan .5 Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In Proceedings bility, of the 2021 ACM Conference on Fairness, Accounta .(623-and Transparency (pp. 610

Wang, T., Chen, X., Hua, G., Wang, X., & Zhang, H. .6 Detecting deepfake videos from multiple .(2020) .representations. arXiv preprint arXiv:2005.05551

Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes .7

IEEE 2019-using inconsistent head poses. In ICASSP 2019 International Conference on Acoustics, Speech and Signal .IEEE .(8265-Processing (ICASSP) (pp. 8261

Lahoti, P., Garimella, V. R. K., & Gionis, A. (2019, .8 representations October). Joint learning for user and item in recommendation. In Proceedings of the 28th ACM International Conference on Information and Knowledge .(2138-Management (pp. 2129

Zellers, R., Bisk, Y., Schwartz, R., & Choi, Y. (2019). .9 nded scale adversarial dataset for grou-SWAG: A large .commonsense inference. arXiv preprint arXiv:1908.11540

Wineburg, S., McGrew, S., Breakstone, J., & Ortega, T. .10

Evaluating information: The cornerstone of civic .(2016)
.online reasoning. Stanford Digital Repository

Wineburg, S., Rapaport, A., ,.Breakstone, J., Smith, M .11
Carle, J., Garland, M., & Saavedra, A. (2019). Students'
civic online reasoning: A national portrait. Educational
.659-Researcher, 48(9), 653

Hao, K. (2021). This paper explains how AI could .12
t to do about it. MIT spread misinformation, and wha
.Technology Review

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & .13
Floridi, L. (2021). The ethics of algorithms: Mapping the
.debate. Big Data & Society, 3(1), 205395171667967

erinsky, A. J., Lazer, D. M., Baum, M. A., Benkler, Y., B .14
Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The
.1096-science of fake news. Science, 359(6380), 1094

Chesney, R., & Citron, D. K. (2018). Deepfakes and the .15
truth -new disinformation war: The coming age of post
.cs. Foreign Aff., 98, 147geopoliti

Gehrmann, S., Strobelt, H., & Rush, A. M. (2021). GLTR: .16
Statistical detection and visualization of generated text. In
Proceedings of the 2021 ACM Conference on Fairness,
.(358-Accountability, and Transparency (pp. 349

Boididou, C., Andreadou, K., Papadopoulos, S., & .17

Drosatos, G. (2021). Verifying the authenticity of images
.77949-using deep learning. IEEE Access, 9, 77938

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, .18
Language models .(2020) .J., Dhariwal, P., ... & Amodei, D
.shot learners. arXiv preprint arXiv:2005.14165-are few

Gehrmann, S., Strobelt, H., & Rush, A. M. (2021). GLTR: .19
Statistical detection and visualization of generated text. In
,Proceedings of the 2021 ACM Conference on Fairness
.(358-Accountability, and Transparency (pp. 349

Major, A., & -Bender, E. M., Gebru, T., McMillan .20
Shmitchell, S. (2021). On the dangers of stochastic
parrots: Can language models be too big?. In Proceedings

ntability, of the 2021 ACM Conference on Fairness, Accou
.(623-and Transparency (pp. 610

تنصل

الآراء الواردة في هذه الأطروحة هي آراء المؤلف ولا تعكس بالضرورة السياسة أو الموقف الرسمي لأي وكالة أو منظمة أو صاحب عمل أو شركة.

يتم توفير أي مواد أو وسائط أو مواقع ويب مشار إليها هنا لأغراض إعلامية فقط. لا يضمن المؤلف دقتها ولا يتحمل أي مسؤولية عن أي معلومات أو تمثيل واردة فيها.

على الرغم من إجراء أبحاث مكثفة من مصادر أكاديمية وتقنية موثوقة، إلا أن هذا العمل قد يحتوي على أخطاء أو سهو غير مقصود. مع استمرار التطورات التكنولوجية والاجتماعية، قد تصبح بعض التفاصيل أو الإحصائيات قديمة.

لا يتم التعبير عن أي توصيات أو موافقات أو علاقات تجارية بين المؤلف وأي طرف ثالث بشكل صريح أو ضمني. هذه الرسالة مخصصة للأغراض التعليمية والبحثية فقط وليس المقصود منها تقديم مشورة قانونية أو مهنية أو غيرها.

يعترف المؤلف بالقيود في المعالجة الكاملة للمشاكل الاجتماعية التقنية المعقدة ويرحب بالمناقشة المحترمة لتعزيز التفاهم الجماعي. بشكل عام، يهدف هذا العمل إلى تعزيز الخطاب العام حول سلامة الذكاء الاصطناعي، وليس اتهام أو مهاجمة أي أفراد أو مجموعات أو منظمات.

المؤلف

ماهر أسعد بكر، كاتب وصحفي وموسيقي سوري. ولد في دمشق عام 1977.